US 20140201741A1

(19) United States

(12) Patent Application Publication

(10) Pub. No.: US 2014/0201741 A1

(43) Pub. Date: Jul. 17, 2014

Govindan et al.

(54) **WORKLOAD INTERFERENCE ESTIMATION AND PERFORMANCE OPTIMIZATION**

(71) Applicant: **Microsoft Corporation**, Redmond, WA (US)

(72) Inventors: **Sriram Govindan**, State College, PA (US); **Jie Liu**, Medina, WA (US); **Aman Kansal**, Issaquah, WA (US)

(73) Assignee: **Microsoft Corporation**, Redmond, WA (US)

(21) Appl. No.: **14/218,985**

(22) Filed: **Mar. 19, 2014**

Related U.S. Application Data

(63) Continuation of application No. 12/843,054, filed on Jul. 26, 2010, now Pat. No. 8,707,300.

Publication Classification

(51) **Int. Cl.**

G06F 9/50 (2006.01)

G06F 9/455 (2006.01)

(52) **U.S. Cl.**

CPC ***G06F 9/5083*** (2013.01); ***G06F 9/45533*** (2013.01)

USPC **718/1**; 718/104

(57) **ABSTRACT**

Architecture that facilitates the estimation of interference among workloads (e.g., virtual machines) due to sharing of a shared resource (e.g., a shared cache of a computer processor), and optimization of a desired performance objective such as power or energy use in the presence of the interference. Estimation is to the extent of interference by characterizing the nature of shared resource usage and its effect on performance. Performance optimization is accomplished using metrics based on the above estimation, or alternatively, an explicit measurement of the interference effects. Methods are employed to estimate interference on the workload's performance with changes in availability of the shared resource or with combinations of other workloads sharing the same resource and allocating workloads to one or more physical computers or resources to workloads such that a desired performance objective is optimized. The methods can include allocating workloads on demand.